New Recognition Mode for A TG Mismatch: The Atomic Structure of a Very Short Patch Repair Endonuclease-DNA Complex

Susan E. Tsutakawa and Kosuke Morikawa,

Departments of Structural Biology, Biomolecular Engineering Research Institute (BERI), 6-2-3 Furuedai, Suita, Osaka 565-0874, Japan

Running title: Vsr/DNA complex

Running title: TG mismatch recognition

Corresponding author Kosuke Morikawa

Departments of Structural Biology

Biomolecular Engineering Research Institute (BERI)

6-2-3 Furuedai, Suita

Osaka 565-0874

Japan

Tel +81-6-6872-8211

Fax +81-6-6872-8219

E-mail: morikawa@beri.co.jp

Introduction The crystal structure determination of T4 endo V pyrimidine photodimer DNA glycosylase provided the first direct view of DNA lesion recognition by a repair enzyme (Vassylyev et al.,1995). Similar damaged DNA recognition modes, involving nucleotide flipping, were observed in various base excision repair enzymes (Parikh et al.,1997, Vassylyev and Morikawa,1997, Mol et al.,1999). The implications of nucleotide flipping raised the new question of how DNA endonucleases other than base excision repair enzymes recognize mismatched base pairs. The *E. coli* very short patch repair (Vsr) endonuclease is a good target to address this question in terms of three-dimensional structures. The Vsr endonuclease is involved in the initial reaction for the repair of mismatched TG base pairs generated through spontaneous deamination of a methylated cytosine. This enzyme recognizes a TG mismatch within the duplex 5'CT(A/T)GG, where the second T forms the mismatch and all of the other bases are in standard Watson-Crick base pairing. It catalyzes the cleavage at the 5' side of the thymine, leaving a 5' phosphate and a 3' hydroxyl at the termini. The crystal structure of a truncated form of this endonuclease was determined at 1.8 Å resolution (Tsutakawa et al.,1999). The protein was found to contain one structural zinc binding module. Unexpectedly, its overall topology resembles members of the type II restriction endonuclease family, although the catalytic center with critical histidines is distinct from those of restriction enzymes. More recently, the crystal structure of Vsr endonuclease in complex with DNA has been determined at 2.3 Å resolution (Tsutakawa et al.,1999). This endonuclease has been found to employ a novel mismatch base pair recognition scheme which does not involve base flipping-out. Extensive interactions between the DNA and

the protein characterize the recognition mechanism, where three aromatic residues intercalate from the major groove into the DNA to strikingly deform the base pair stacking. An amino terminal α-helix is accommodated into the expanded minor groove so that the amino acid side chains make additional contacts with the DNA duplexes. With the presence of a cleaved DNA intermediate in the active center, the structure of the Vsr/DNA complex provides detailed insights into the catalytic mechanism for endonuclease activity.

Recognition of Damaged DNA DNA damage recognition and repair are critical cell functions to ensure transcriptional viability and genomic stability. Environmental and internal stresses can result in a myriad of base damages, many of which the cell must be able to recognize and respond. A battery of enzymes have evolved for recognition and resolution of abnormal bases: photolesions, alkylated bases, abasic sites, breaks in the phosphodiester backbone, etc. Structures of DNA repair enzymes in complex with damaged DNA have revealed nucleotide-flipping as the common means of recognizing and resolving DNA damage. The greater accessibility of the base provides a catalytic rationale for nucleotide flipping for the six glycosylases: T4 endoV (Vassylyev et al.,1995); human uracil-DNA glycosylase (Parikh et al.,1998, Parikh et al.,2000, Slupphaug et al.,1996); bacterial mismatch-specific uracil-DNA glycosylase (MUG) (Barrett et al.,1998, Barrett et al.,1999); human 3-methyladenine DNA glycosylase (Lau et al.,1998); *E. coli* AlkA (Hollis et al.,2000); and 8-oxoguanine DNA glycosylase (hOGG1) (Bruner et al.,2000), all use nucleotide flipping either of the damaged base and/or the base opposite for recognition and catalysis. For the apurinic/apyrimidinic

endonucleases, flipping out abasic sites into a shallow niche which would exclude any bases is important for substrate recognition in bacterial EndoIV (Hosfield et al.,1999) and human APE1 (Mol et al.,2000). It is also thought that nucleotide flipping helps in the choreography of base excision repair, with the damaged DNA getting passed from one protein to another in the multi-enzyme repair pathway (Mol et al.,1999, Parikh et al.,1999, Wilson and Kunkel,2000).

However, damage that results in mismatched DNA poses a distinct problem, since recognition must be of natural bases in an unnatural context. It was not clear, especially in light of the work done for damaged DNA repair systems, whether mismatch recognition would also proceed through a flipping out mechanism.

Very Short Patch Repair In *E. coli,* there are three systems that have developed to recognize mismatched DNA: long patch repair, very short patch repair (VSR), and the MutY system. Long patch repair, initiated by the MutHLS complex, is the most ubiquitous and is responsible for repair of most mismatches. MutS recognizes the mismatch, and then, in complex with MutL, travels down the DNA in search of a methylation site. To prevent mismatches that occur during replication to propagate to future generations, methylation provides the cell a means for preferentially repairing younger, not yet methylated strands rather than older, parental strands. MutH cuts opposite to a methylated base, then the DNA is excised back to the original mismatch and the strand is repolymerized with the correct bases.

However, a problem arises when the damage occurs at the actual methylation site (Fig. 1). Methylated cytosines are vulnerable to spontaneous deamination and hydrolysis, resulting in

a thymine mismatched to a guanine. The MutHLS system would preferentially repair the non-methylated strand, resulting in the mutation of guanine to adenine. This would also lead to the eventual loss of recognition sites for the methyltransferase. The VSR system, first identified by Margaret Lieb in a series of elegant genetic studies of λ phage, ensures the correct repair of damage occurring at cytosine methylation sites (Lieb,1985).

Vsr, the central protein in the VSR system, is responsible for recognition of the TG mismatch and for cleavage of the phosphodiester backbone to initiate repair. The tangential and partially overlapping gene arrangement of *vsr* and *dcm* appears to ensure equivalent levels of expression from a single transcript (Lieb,1987, Zell and Fritz,1987, Sohail et al.,1990, Dar and Bhagwat,1993). Hennecke and coworkers were the first to identify Vsr as an endonuclease and determined that Vsr cleaves the phosphodiester backbone on the 5' side of the mismatched thymine leaving 5' phosphate and 3' hydroxyl termini (1991). Magnesium is absolutely required for this activity (Hennecke et al.,1991, Tsutakawa et al.,1999). Vsr is specific for TG mismatches within the Vsr substrate sequence $^{5'}$CT(A/T)AGG$^{3'}$/$^{3'}$GG(T/A)CC$^{5'}$ (Lieb,1985, Hennecke et al.,1991), which corresponds to the recognition sequence for Dcm methyltransferase (Hattman et al.,1973). The first, second, and fourth base pairs are absolutely required for Vsr recognition (Glasner et al.,1995). MutL and MutS were also found to be required for efficient repair of the TG mismatch, although their role is unclear (Lieb,1987, Zell and Fritz,1987, Lieb and Rehmat,1995). MutL has been shown to enhance binding of Vsr to DNA, and it has been proposed that MutLS may present the TG mismatch to Vsr in a loop context (Drotschmann et

al.,1998, Rasmussen et al.,1988). Only five residues need to be removed for repair of the mismatch (Lieb et al.,1986), presumably replaced by DNA Polymerase I, which is required for VSR repair (Dzidic and Radman,1989). Polymerase activity by the DNA polymerase I and subsequent ligation would complete repair of the site.

Structural studies of Vsr endonuclease Vsr is a small 18 kD protein and is conserved in *E. coli, Xanthomonas oryzae, Nocardia aerocolonigenes,* and *Haemophilus parainfluenzae* (Hanck et al.,1989, Choi and Leach,1994, Kulakauskas et al.,1994, Taron et al.,1995). The structure of an N-terminal truncated form of the *E. coli* protein was determined to 1.8 Å resolution (Tsutakawa et al.,1999) and revealed a central β-sheet, framed by α-helices (Fig. 2). A structural zinc site, coordinated by three residues on one loop (Cys-66, His-71, and Cys-73) and by one distal residue (Cys-117), appears to be integral for stabilizing the overall structure of this relatively small protein. The zinc site is similar only to one found in I-PpoI homing endonuclease, the first of many structural similarities with this protein.

The overall fold of Vsr resembles members of the type II restriction endonuclease family. Proteins in this family are characterized by a conserved overall topology with a central beta sheet subdivided into two short strands followed by a set of longer strands, and their active site falls into the cleft formed by the short and long strands of a central β-sheet, with the first conserved acidic residue of the catalytic motif, $DX_{(6-30)}$(D/E)XK jutting out from the end of the last short β-strand and the remaining two conserved residues on the length of the adjacent long strand

(Aggarwal,1995, Pingoud and Jeltsch,1997). These residues are virtually superimposable. The first two acidic residues are involved in the catalytic metal coordination and the last residue, which is actually the least conserved of the three, has been proposed to activate the attacking water in the phosphodiesterase mechanism. Asp51 in Vsr easily superimposes onto the first aspartate of the type II restriction enzyme catalytic motif (Fig. 3), and alanine scanning confirmed its critical role in catalysis. In contrast, the remainder of the catalytic motif was not as well-conserved in Vsr. Phe62 is in the same position as the second catalytic residues and would obviously not be involved in binding metal, although it is invariantly conserved in Vsr-related proteins, In the third position which is admittedly the least conserved, His64 would be the first histidine to occupy this position. However, an alanine mutant, which retained partial endonuclease activity, demonstrated that His64 is not an essential catalytic residue. H69A, which is nearby but does not superimpose onto the catalytic motif, was completely inactive. The presence and significance of these histidines represent significant discrepancies between Vsr and type II restriction enzyme family members.

Active site with DNA intermediate and two magnesium-water clusters. The 2.3 Å structure of Vsr in complex with DNA (Fig. 2) clarified the distinctive qualities of the endonuclease (Tsutakawa et al.,1999). The protein was crystallized in the presence of magnesium and intact dsDNA containing a TG mismatch within the five basepair conserved sequence. During the three weeks that it took to nucleate and form diffraction quality crystals, the phosphodiester backbone had been cleaved, providing a glimpse at the active site containing

a cleaved DNA intermediate. Fritz and coworkers (1991) found that Vsr endonuclease cleaves on the 5' side of the mismatched thymine base and leaves 5' phosphate and 3' hydroxyl termini, and this result was clearly confirmed by electron density maps after simulated annealed omit maps. The DNA termini were coordinated with two magnesiums: the 5' phosphate is coordinated to both magnesiums, while the 3' ribose oxygen coordinates only one metal (Fig. 4). The observed 2.3 Å distance between the 3' termini and the metal would preclude the final protonation of the oxygen and identifies the DNA as a cleaved intermediate.

In the only residue-specific interaction, Asp51 directly coordinates the two catalytic metals (Fig. 4). The backbone carbonyl of Thr63 provides a second direct protein-metal contact, with the remainder of the coordination through the water cluster surrounding the magnesium ions, initially observed as octahedrally-shaped electron densities. The importance of these waters is evidenced in the significant reductions in activity upon alanine substitutions of Asp25, a residue which is involved in water-mediated coordination of both magnesiums, and of His64, which is binding to one of these magnesium-water clusters (Tsutakawa et al.,1999).

His69, shown to be an essential catalytic residue by alanine substitution (Tsutakawa et al.,1999), and one of the magnesiums are both clearly interacting with one of the cleaved 5' phosphate oxygens, revealing a potential mechanism for cleavage. This oxygen is likely to have been part of the attacking water, with the magnesium acting as Lewis acid and with His69, part of a His-Asp pair, accepting the proton. The significance of the His-Asp pair is underscored by the decrease in catalytic activity upon substitution of that aspartate, Asp97, by alanine. The

geometry of this water would also agree with an in-line attack on the phosphodiester bond. Then both magnesium ions would stabilize the pentacoordinate intermediate, analogous to the two-metal mechanism first described for DNA polymerase I (Beese and Steitz,1991). There is no clear protein sidechain candidate, either by geometry or by sequence conservation, for the last chemical step in the reaction, the reduction of the leaving oxygen. One of the waters surrounding the magnesiums could easily play that role.

In light of the similarity in overall topology with type II restriction endonucleases, the active site and mechanism is surprisingly more in line with those from I-PpoI homing endonuclease. The water-magnesium clusters, found in the active site of Vsr, were initially observed for Serratia endonuclease (Miller et al.,1999) and have also been found in the related I-PpoI homing endonuclease (Flick et al.,1998), the latter whose structure has been solved in the presence of substrate and product DNA (Flick et al.,1998, Galburt et al.,1999). I-PpoI homing endonuclease, with a single magnesium-water cluster, also uses a histidine to activate and orient the attacking water (Flick et al.,1998, Galburt et al.,1999). A backbone carboxylate with the proper distance and orientation for a hydrogen bond to the histidine appears to mimic the his-asp interaction observed for His69-Asp97 in Vsr. The observation that the sugars on either side of the cleavage are $C^{3'}$ endo for both Vsr and I-PpoI endonucleases is particularly intriguing in that sugar conformation may promote catalysis by orienting the phosphate for catalysis and that this sugar conformation is retained in the product complexes.

As found for type II restriction endonucleases, similar topology does not necessarily translate into identical mechanisms for catalysis (Galburt and Stoddard,2000), and Vsr is not an exception. Using a two-magnesium-water-cluster based mechanism and a His-Asp pair to activate the attacking water, the details of endonuclease catalysis are unique to Vsr.

Protein/DNA interaction. The most striking feature of the Vsr/DNA complex is the mode of TG mismatch recognition. The N-terminal helix of Vsr clasps the DNA down onto the core region of the protein (Fig. 2). The strength of the interaction is underscored by the fact that a three base pair long sequence, generated after cleavage, remained during the three week long crystallization at 20°C. With all residues maintained as base pairs including the TG mismatch as a wobble base pair, Vsr also makes a remarkable three residue intercalation into the DNA, on the 5' side of the mismatched thymine (Fig. 5). This result is in contrast to the nucleotide flipping observed for all other specific damage-recognizing enzymes to date.

The N-terminus is apparently flexible and was easily cleaved by limited proteolysis initially at Arg15, with subsequent cleavage at Arg20 (unpublished data). The directed nature of the proteolysis was explained by the finding that Arg15 lies at the start of a loop region between the first and second helix, which was initially predicted to be a single helix. We envision that this region acts as a hinge, that can snap shut upon DNA binding.

The DNA is pierced through from the major groove side by three invariantly conserved residues, Phe67, Trp68, and Trp86, which stack against the bases and sugars of the DNA on one side of the TG mismatch. Trp68 directly stacks against the 5' side of the mismatched thymine

base, while Phe67 interacts with bases of the AT basepair at the center of the recognition sequence and Trp86 with the two sugars opposite of the thymine base. It is notable the Trp68, which has the largest interface with the mismatched thymine, is the sidechain most strikingly different between the DNA-bound and apo-forms. There is an approximately 110° rotation of the indole ring out of the main plane of the protein core. Met14 and Ile17 complete the piercing of the DNA by coming in from the minor groove side (Fig. 6). It is interesting to postulate that the reasons as to why this N-terminal region of Vsr is seemingly the most flexible part of the whole protein, involve possible mechanisms of damage recognition. The intercalation leaves only enough room for phosphates to pass.

Vsr is not the first protein found to intercalate into DNA duplexes. However, it is distinct from the others, which surprisingly all fall into the same category (reviewed in (Werner et al.,1996). Generally, they are involved in transcriptional regulation or DNA stabilization, and none are enzymes. One to four protein residues form a small wedge and partially insert from the minor groove side into the DNA. DNA stacking is disrupted, resulting in an A-type DNA with an enlarged minor groove and a compressed major groove. For example, in the TATA box DNA/protein complex, two sets of two phenylalanine residues intercalate into the DNA and assist in bending it by approximately 70° (Fig. 6). Vsr, on the other hand, goes completely through the DNA, with Phe67, Trp68, and Trp86 stacking with the nucleotides from the major groove side and Met14 and Ile17 coming in from the minor groove side. As a result, both the major and minor grooves are significantly widened, underscored by the unprecedented presence

of a helix deep in the minor groove. Although the increase in rise, 6 Å for Vsr, is greater than the other intercalating proteins, the overall bend is only about 40°.

TG mismatch recognition Vsr recognizes the TG mismatch in an intact base pair, in direct contrast to other complexes which all recognize damaged DNA by nucleotide flipping (Fig. 5 and 7). Even two mismatch recognizing proteins, MutY (Guan et al.,1998) and uracil DNA-glycosylases (Slupphaug et al.,1996, Barrett et al.,1998, Parikh et al.,1998) flip out uridine from a base pair with guanine. Classical wobble base pair hydrogen bonding between O6 and N1 of the guanine and O2 and N3 of the thymine, respectively, is identical to that seen in structural studies of DNA containing TG mismatches (Kneale et al.,1985, Hunter et al.,1986, Hunter et al.,1987, Allawi and SantaLucia,1998). In addition, the water bridging the N2 of the guanine and the O2 of the thymine is retained in the protein-bound form, although the major groove waters now appear to be displaced.

The number of direct hydrogen bonds with the protein and the two mismatched bases are surprisingly few in number (Fig. 7). Lys89 and the main chain carbonyl of Met14 specifically recognize the guanine base through O6 and N2, respectively, while Asn93 is the only residue to differentiate thymine from cytosine by hydrogen binding to O4. The paucity of direct interactions is also found for the entire specific recognition sequence, $^{5'}$CTAGG$^{3'}$/$^{3'}$GGTCC$^{5'}$, with an average of one direct interaction per base. However, there is an extensively intertwined network of water-mediated interactions, where one residue makes multiple direct and indirect contacts to one to four bases. There is almost an equivalent number of contacts with the

phosphate backbone. The large percentage (40%) of these direct protein-phosphate interactions which stem from the amide backbone, suggests that stabilization of the phosphate backbone through a less flexible interface is important for Vsr function.

Along these lines, the protein surface is extremely complementary to the DNA, especially at the position of the TG mismatch. Originally observed in structural studies with TG mismatch-containing DNA (Hunter et al.,1986), the thymine base shifts into the major groove in order to make the wobble base pair hydrogen bonding, and the surface curvature of Vsr follows this shift (Fig. 5 and 7). The presence of cytosine, which is not a substrate for endonuclease activity (Hennecke et al.,1991), would shift the base and the phosphate backbone up out of the active site cleft and push the N-terminus up. Since Glu25 is involved in coordination of both catalytic magnesium-water clusters, disruption of the N-terminus could further decrease activity, as seen for the N-terminal deletion mutants (Tsutakawa et al.,1999).

Unlike DNA repair enzymes where nucleotide-flipping is required to properly place the scissile bonds into the active site, to allow catalytic residues access, or to optimize electron orbital orientation in the substrate; intercalation surprisingly occurs on the side of the TG mismatch opposite to the cleavage and therefore does not directly position the substrate for catalysis. Instead, we propose that intercalation is playing a key role in TG mismatch recognition. One of the few DNA parameters affected by the presence of a TG wobble base pair in B-form DNA is a significant 1 Å increase in rise on one side of the TG base pair (Hunter et al.,1987). This increase in rise, which is the distance from one base pair to the next, can be

attributed to changes in base stacking. Shifting of the thymine base into the major groove enhances the base pair overlap on one side while disrupting stacking on the other side (Fig. 7). Intercalation in the Vsr/DNA complex occurs on the side of the TG mismatch with the disrupted stacking. Thus, we envision that Vsr not only uses specific hydrogen bonding and steric complementarity for recognition of the TG wobble base pair, but in fact also differentiates the mismatch from normal DNA by searching for disrupted base pair stacking.

Vsr is an important exception to many rules: to the catalytic motif rule of the type II restriction endonuclease family, to the intercalation rule of DNA binding enzymes, and, most important for understanding DNA repair, to the nucleotide flipping rule for recognition and resolution of DNA damage. Thus far, all DNA repair enzymes, whose DNA complex structures have been determined, have flipped out their bases. For the DNA glycosylases, the enzymes must make access to the base-sugar bond, and they use nucleotide flipping to either flip out the damaged base or flip out the base opposite to the damage to make room for catalytic residues to move in. In the case of APE1 and endoIV, which like Vsr cleave the exposed phosphodiester backbone, flipping out provides an easy means to recognize abasic sites. In the case of Vsr, the accessibility of the phosphodiester backbone and the ease of recognizing the TG wobble base pair precludes the necessity of nucleoside flipping. Therefore, the distinct mismatch recognition mechanism of VSR underscores the idea that specific functional requirements of each DNA repair enzyme, either in terms of substrate recognition or of catalysis, dictate the mode of DNA recognition.

Acknowledgements

We are grateful to Prof. M. Nakagawa and Drs. T. Kawate and H. Jingami for their contribution to this study.

References

Aggarwal, A. K. 1995. Structure and function of restriction endonucleases. *Curr Opin Struct Biol*. **5:**11-19.

Allawi, H. T., and Santalucia, J., Jr. 1998. NMR solution structure of a DNA dodecamer containing single g*t mismatches. *Nucleic Acids Res*. **26:**4925-4934.

Barrett, T. E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J., and Pearl, L. H. 1998. Crystal structure of a G:T/U mismatch-specific DNA glycosylase: Mismatch recognition by complementary-strand interactions. *Cell*. **92:**117-129.

Barrett, T. E., Scharer, O. D., Savva, R., Brown, T., Jiricny, J., Verdine, G. L., and Pearl, L. H. 1999. Crystal structure of a thwarted mismatch glycosylase DNA repair complex. *EMBO J*. **18:**6599-6609.

Beese, L. S., and Steitz, T. A. 1991. Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: A two metal ion mechanism. *EMBO J*. **10:**25-33.

Bruner, S. D., Norman, D. P., and Verdine, G. L. 2000. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature*. **403:**859-866.

Choi, S. H., and Leach, J. E. 1994. Identification of the *Xorii* methyltransferase gene and a vsr homolog from *Xanthomonas oryzae* pv. Oryzae. *Mol Gen Genet*. **244:**383-390.

Dar, M., and Bhagwat, A. S. 1993. Mechanism of expression of DNA repair gene *vsr*, an *Escheria coli* gene that overlaps the DNA cytosine methylase gene, *dcm*. *Mol Microbiol*. **9:**823-833.

Drotschmann, K., Aronshtam, A., Fritz, H. J., and Marinus, M. G. 1998. The *Escherichia coli* MutL protein stimulates binding of Vsr and MutS to heteroduplex DNA. *Nucleic Acids Res*. **26:**948-953.

Dzidic, S., and Radman, M. 1989. Genetic requirements for hyper-recombination by very short patch mismatch repair: Involvement of *Escherichia coli* DNA polymerase I. *Mol Gen Genet*. **217:**254-256.

Flick, K. E., Jurica, M. S., Monnat, R. J., Jr., and Stoddard, B. L. 1998. DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI *Nature*. **394:**96-101.

Galburt, E. A., Chevalier, B., Tang, W., Jurica, M. S., Flick, K. E., Monnat, R. J., Jr., and Stoddard, B. L. 1999. A novel endonuclease mechanism directly visualized for I-PpoI. *Nat Struct Biol*. **6:**1096-1099.

Galburt, E. A., and Stoddard, B. L. 2000. Restriction endonucleases: One of these things is not like the others. *Nat Struct Biol*. **7:**89-91.

Glasner, W., Merkl, R., Schellenberger, V., and Fritz, H. J. 1995. Substrate preferences of Vsr DNA mismatch endonuclease and their consequences for the evolution of the *Escherichia coli* K-12 genome. *J Mol Biol*. **245:**1-7.

Guan, Y., Manuel, R. C., Arvai, A. S., Parikh, S. S., Mol, C. D., Miller, J. H., Lloyd, S., and Tainer, J. A. 1998. MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily. *Nat Struct Biol.* **5:**1058-1064.

Hanck, T., Gerwin, N., and Fritz, H. J. 1989. Nucleotide sequence of the *dcm* locus of *Escherichia coli* K12 *Nucleic Acids Res*. **17:**5844.

Hattman, S., Schlagman, S., and Cousens, L. 1973. Isolation of a mutant of *Escherichia coli* defective in cytosine- specific deoxyribonucleic acid methylase activity and in partial protection of bacteriophage lambda against restriction by cells containing the N-3 drug-resistance factor *J Bacteriol*. **115:**1103-1107.

Hennecke, F., Kolmar, H., Brundl, K., and Fritz, H. J. 1991. The vsr gene product of *E. coli* K-12 is a strand- and sequence-specific DNA mismatch endonuclease. *Nature*. **353:**776-778.

Hollis, T., Ichikawa, Y., and Ellenberger, T. 2000. DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, *Escherichia coli* AlkA. *EMBO J*. **19:**758-766.

Hosfield, D. J., Guan, Y., Haas, B. J., Cunningham, R. P., and Tainer, J. A. 1999. Structure of the DNA repair enzyme endonuclease IV and its DNA complex: Double-nucleotide flipping at abasic sites and three-metal-ion catalysis. *Cell*. **98:**397-408.

Hunter, W. N., Brown, T., Kneale, G., Anand, N. N., Rabinovich, D., and Kennard, O. 1987. The structure of guanosine-thymidine mismatches in B-DNA at 2.5-Å resolution. *J Biol Chem*. **262:**9962-9970.

Hunter, W. N., Kneale, G., Brown, T., Rabinovich, D., and Kennard, O. 1986. Refined crystal structure of an octanucleotide duplex with G . T mismatched base-pairs. *J Mol Biol*. **190:**605-618.

Kim, Y. C., Grable, J. C., Love, R., Greene, P. J., and Rosenberg, J. M. 1990. Refinement of EcoRI endonuclease crystal structure: A revised protein chain tracing. *Science*. **249:**1307-1309.

Kneale, G., Brown, T., Kennard, O., and Rabinovich, D. 1985. G . T base-pairs in a DNA helix: The crystal structure of d(g-g-g-g-t-c- c-c). *J Mol Biol*. **186:**805-814.

Kulakauskas, S., Barsomian, J. M., Lubys, A., Roberts, R. J., and Wilson, G. G. 1994. Organization and sequence of the HpaII restriction-modification system and adjacent genes. *Gene*. **142:**9-15.

Lau, A. Y., Scharer, O. D., Samson, L., Verdine, G. L., and Ellenberger, T. 1998. Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: Mechanisms for nucleotide flipping and base excision. *Cell*. **95:**249-258.

Lieb, M. 1985. Recombination in the lambda repressor gene: Evidence that very short patch (vsp) mismatch correction restores a specific sequence. *Mol Gen Genet*. **199:**465-470.

Lieb, M. 1987. Bacterial genes mutl, muts, and dcm participate in repair of mismatches at 5-methylcytosine sites. *J Bacteriol*. **169:**5241-5246.

Lieb, M., Allen, E., and Read, D. 1986. Very short patch mismatch repair in phage lambda: Repair sites and length of repair tracts. *Genetics*. **114:**1041-1060.

Lieb, M., and Rehmat, S. 1995. Very short patch repair of T:G mismatches *in vivo*: Importance of context and accessory proteins. *J Bacteriol*. **177:**660-666.

Miller, M. D., Cai, J., and Krause, K. L. 1999. The active site of Serratia endonuclease contains a conserved magnesium-water cluster *J Mol Biol*. **288:**975-987.

Mol, C. D., Izumi, T., Mitra, S., and Tainer, J. A. 2000. DNA-bound structures and mutants reveal abasic DNA binding by Ape1 DNA repair and coordination. *Nature*. **403:**451-456.

Mol, C. D., Parikh, S. S., Putnam, C. D., Lo, T. P., and Tainer, J. A. 1999. DNA repair mechanisms for the recognition and removal of damaged DNA bases. *Annu Rev Biophys Biomol Struct*. **28:**101-128.

Parikh, S. S., Mol, C. D., Hosfield, D. J., and Tainer, J. A. 1999. Envisioning the molecular choreography of DNA base excision repair. *Curr Opin Struct Biol*. **9:**37-47.

Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E., and Tainer, J. A. 1998. Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J*. **17:**5214-5226.

Parikh, S. S., Mol, C. D., and Tainer, J. A. 1997. Base excision repair enzyme family portrait: Integrating the structure and chemistry of an entire DNA repair pathway. *Structure*. **5:**1543-1550.

Parikh, S. S., Walcher, G., Jones, G. D., Slupphaug, G., Krokan, H. E., Blackburn, G. M., and Tainer, J. A. 2000. Uracil-DNA glycosylase-DNA substrate and product structures: Conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proc Natl Acad Sci U S A*. **97:**5083-5088.

Pingoud, A., and Jeltsch, A. 1997. Recognition and cleavage of DNA by type-II restriction endonucleases. *Eur J Biochem*. **246:**1-22.

Rasmussen, L. J., Samson, L., and Marinus, M. G. 1988. Dam-directed DNA mismatch repair. In *DNA damage and repair* J. A. Nickoloff, and M. Hoekstra (eds) p. 205-228. Humana Press, Totowa, NJ.

Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E., and Tainer, J. A. 1996. A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature*. **384:**87-92.

Sohail, A., Lieb, M., Dar, M., and Bhagwat, A. S. 1990. A gene required for very short patch repair in *Escherichia coli* is adjacent to the DNA cytosine methylase gene. *J Bacteriol*. **172:**4214-4221.

Taron, C. H., Van Cott, E. M., Wilson, G. G., Moran, L. S., Slatko, B. E., Hornstra, L. J., Benner, J. S., Kucera, R. B., and Guthrie, E. P. 1995. Cloning and expression of the NaeI restriction endonuclease-encoding gene and sequence analysis of the NaeI restriction-modification system *Gene*. **155:**19-25.

Tsutakawa, S. E., Jingami, H., and Morikawa, K. 1999. Recognition of a TG mismatch: The crystal structure of very short patch repair endonuclease in complex with a DNA duplex. *Cell*. **99:**615-623.

Tsutakawa, S. E., Muto, T., Kawate, T., Jingami, H., Kunishima, N., Ariyoshi, M., Kohda, D., Nakagawa, M., and Morikawa, K. 1999. Crystallographic and functional studies of very short patch repair endonuclease. *Mol Cell*. **3:**621-628.

Vassylyev, D. G., Kashiwagi, T., Mikami, Y., Ariyoshi, M., Iwai, S., Ohtsuka, E., and Morikawa, K. 1995. Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: Structural basis for damaged DNA recognition. *Cell*. **83:**773-782.

Vassylyev, D. G., and Morikawa, K. 1997. DNA-repair enzymes. *Curr Opin Struct Biol*. **7:**103-109.

Wah, D. A., Hirsch, J. A., Dorner, L. F., Schildkraut, I., and Aggarwal, A. K. 1997. Structure of the multimodular endonuclease FokI bound to DNA. *Nature*. **388:**97-100.

Werner, M. H., Gronenborn, A. M., and Clore, G. M. 1996. Intercalation, DNA kinking, and the control of transcription. *Science*. **271:**778-784.

Wilson, S. H., and Kunkel, T. A. 2000. Passing the baton in base excision repair. *Nat Struct Biol*. **7:**176-178.

Zell, R., and Fritz, H. J. 1987. DNA mismatch-repair in *Escherichia coli* counteracting the hydrolytic deamination of 5-methyl-cytosine residues. *EMBO J*. **6:**1809-1815.

Figure Legends

Figure 1. Vsr system is responsible for repair of damaged methylated cytosines.

Figure 2. Crystal structures of Vsr endonuclease: the truncated form (left, 1vsr.pdb) and the full-length in complex with DNA (right, 1cwo.pdb). The proteins are displayed as ribbon models, and the DNA is colored blue and red. The mismatched thymine base is colored red. The zinc atom is displayed as a magenta sphere, and residues coordinating the zinc are displayed in the left figure only. The two magnesium atoms are colored pink (right only).

Figure 3. Superimposition of the Vsr active site (blue) with two type II restriction endonucleases, EcoRI (magenta) and FokI (red). The coordinate files for Vsr, EcoRI, and FokI are 1vsr.pdb, 1eri.pdb (Kim et al.,1990), and 1fok.pdb (Wah et al.,1997).

Figure 4. The active site of Vsr/DNA complex. The F_O-F_C simulated anneal omit map where the metal-water clusters have been omitted, has been contoured at 5 and 11 σ. Magnesium and water atoms are depicted respectively as red and white spheres. Atoms are colored according to element; carbon in green, nitrogen in blue, phosphorus in yellow, and oxygen in red. Interactions within metal or water coordination limits are shown as black dashed lines. The cleaved scissile bond is depicted in red.

Figure 5. Distinct DNA binding modes of Vsr and Uracil DNA glycosylase (Slupphaug, 1996). Surface representation of the VSR protein, with the N-terminus removed for clarity, and liquorice model of the DNA shows the intercalation of three aromatic residues into the completely paired duplex DNA (upper panel). On the other hand, UDG glycosylase flips out the

uracil (lower panel). DNA atoms are colored according to element, with carbon in white, oxygen in red, nitrogen in blue, and phosphorus in yellow.

Figure 6. Intercalation of Vsr is distinct from other intercalating proteins. Vsr (upper panel) completely penetrates the DNA helix, in contrast to the TBP complex (middle panel) panel) where the bases still are partially in contact with each other. The DNA is colored in blue and gold, and the protein residues are colored according to element; carbon in green, nitrogen in blue, sulfur in yellow, and oxygen in red. In the lowest panel, the minor groove of the DNA is displayed with the atoms colored as in Fig 4.

Figure 7. TG mismatch recognition. In the upper panel, the $2F_O$-F_C simulated anneal omit map where the mismatched base pairs were omitted, has been contoured at 1.5 σ. Wobble base pair hydrogen bonding is depicted as red dashed lines, while interaction with surrounding residues and water atoms are depicted in black. Atoms are colored as described in Figure 6. The lower panel shows the effect of TG wobble base pairing (emphasized with the black arrows) on the base pair stacking. Coordinates are from NDB entry BDL009 (Hunter, 1987). Base pairs are colored differently for clarity.

Figure 1

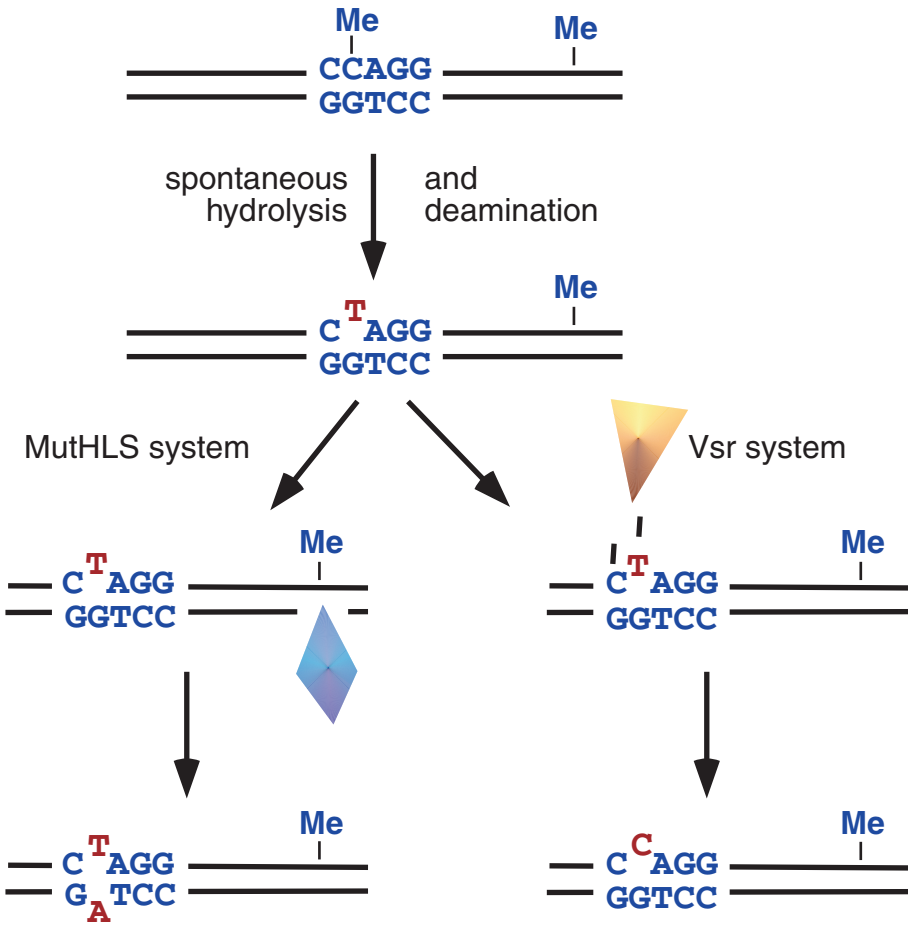
Me
Me
CCAGG
GGTCC
spontaneous
hydrolysis
and
deamination
Me
C T AGG
GGTCC
MutHLS system
Vsr system
Me
C T AGG
GGTCC
Me
C T AGG
GGTCC
Me
C T AGG
G A TCC
Me
C C AGG
GGTCC

Figure 2

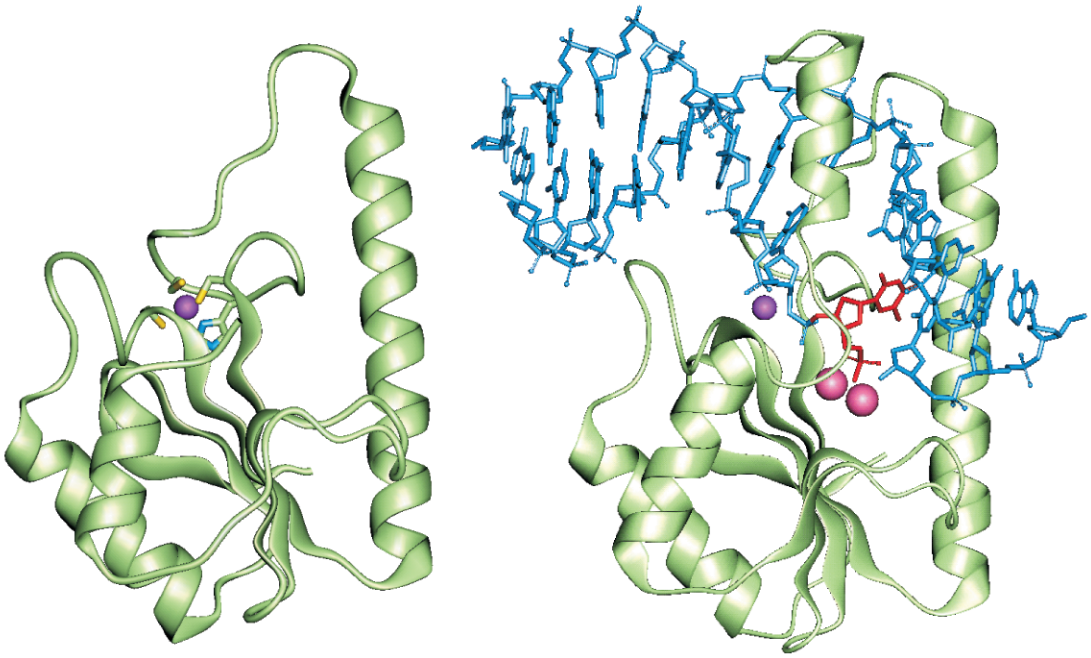

Figure 3

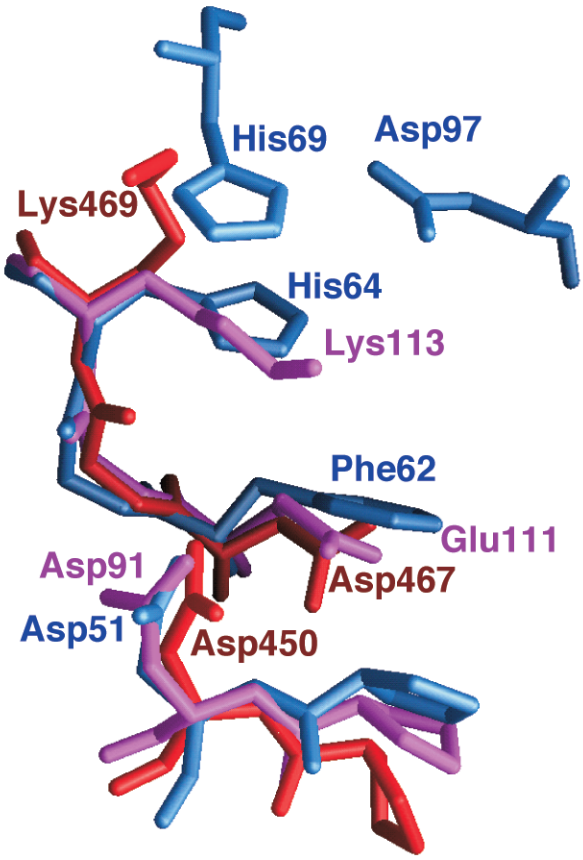
His69
Asp97
Lys469
His64
Lys113
Phe62
Glu111
Asp91
Asp467
Asp51
Asp450

Figure 4

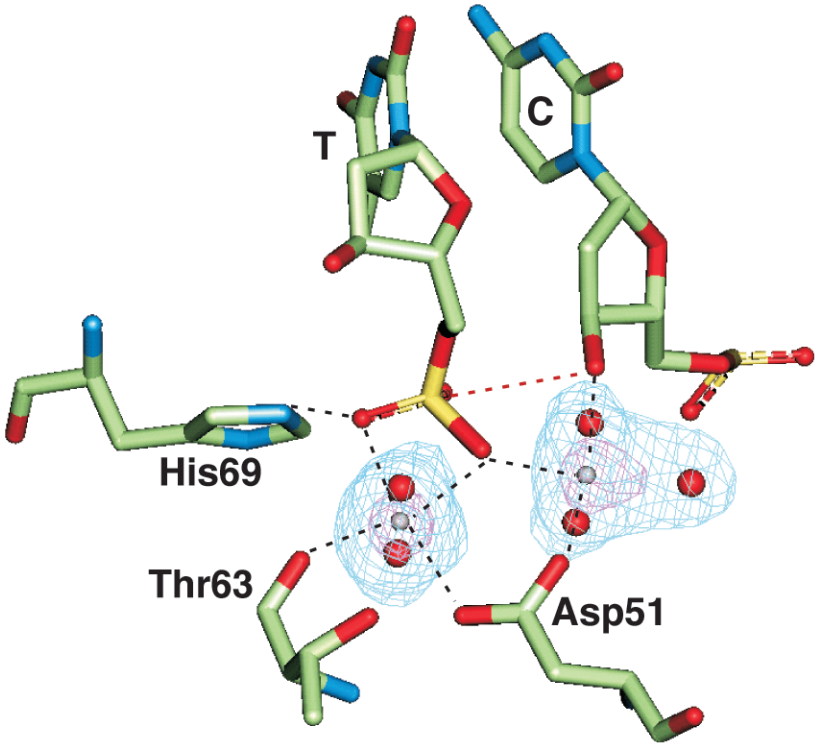

Figure 5

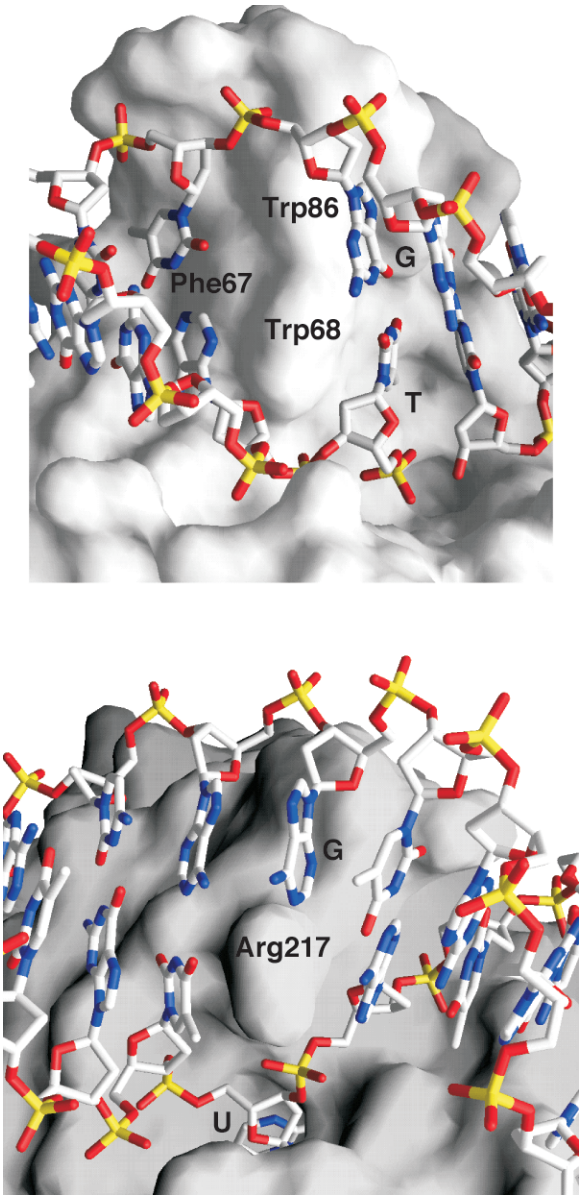

Figure 6

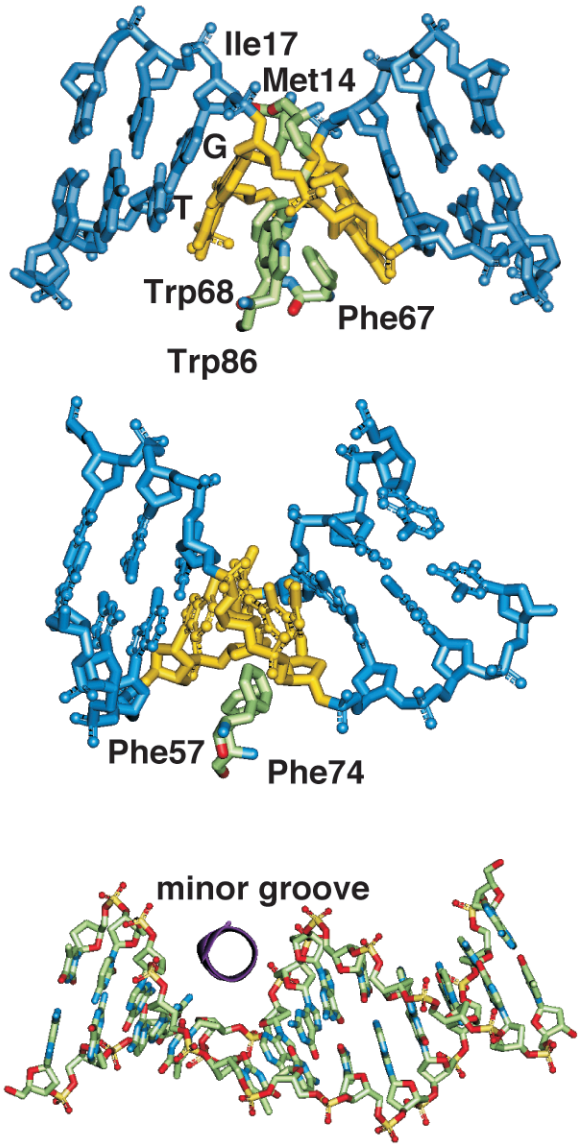

Figure 7